REPORT DOCUMENTATION PAGE

4001951

Form Approved
OMB No 0704-0188

(2)

AD-A270 730

...average ... per response, including the time for reviewing instructions, searching existing data sources ... Send comments regarding this burden estimate or any other aspect of this ... to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson ... Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

...ATE	3. REPORT TYPE AND DATES COVERED
	FINAL 01 Mar 90 TO 31 Jul 93

HUMAN-MACHINE INTERFACES

5. FUNDING NUMBERS

AFOSR-90-0200
61102F
2313
CS

6. AUTHOR(S)

DR NATHANIEL DURLACH

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Research Laboratory of Electronics
Massachusetts Institute of Technology
77 Massachusetts Ave
Cambridge, MA 02139

8. PERFORMING ORGANIZATION REPORT NUMBER

AFOSR-TR- 93 0769

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/NL
110 Duncan Avenue, Suite B115
Bolling AFB DC 20332-0001

LtCol Collins

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

DTIC ELECTE OCT 18 1993 S E D

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release;
distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

The normal human auditory system suffers from many deficiencies in its ability to localize sound sources in space. Not only is it generally poor at determining the elevation and distance of a sound source, but in certain cases it is relatively poor at determining the azimuth of the source. The research discussed in this report is concerned with the development and evaluation of systems that result in improved localization, i.e., in supernormal auditory localization, by altering the localization cues that are available to the listener. Although such enhanced performance should be of value in essentially all systems that make use of auditory localization for conveying information to the human user, the application area of primary interest in this proposal is that of human-machine interfaces for teleoperator and virtual-environment systems. In general, localization performance can be summarized in terms of (1) resolution and (2) response bias. Resolution refers to the ability to detect small changes in the spatial position of a sound source and to separate out multiple sources located at different positions, as well as to the amount of information transfer that can be achieved in the identification of source position. Response bias refers to the average differences between perceived source position (as measured by the mean of the listener's objective responses) and the actual source position.

14. SUBJECT TERMS

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
(U)	(U)	(U)	(U)

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std. Z39-18
298-102

16 SEP 1993

Final Report
to
the Air Force Office of Scientific Research

from research grant

AFOSR 90-0200

Super Auditory Localization for Improved Human-Machine Interfaces

Principal Investigators:
Nathaniel I. Durlach
Richard M. Held

SEP 1993

Abstract

The normal human auditory system suffers from many deficiencies in its ability to localize sound sources in space. Not only is it generally poor at determining the elevation and distance of a sound source, but in certain cases it is relatively poor at determining the azimuth of the source.

The research discussed in this report is concerned with the development and evaluation of systems that result in improved localization, i.e., in supernormal auditory localization, by altering the localization cues that are available to the listener. Although such enhanced performance should be of value in essentially all systems that make use of auditory localization for conveying information to the human user, the application area of primary interest in this proposal is that of human-machine interfaces for teleoperator and virtual-environment systems.

In general, localization performance can be summarized in terms of (1) resolution and (2) response bias. Resolution refers to the ability to detect small changes in the spatial position of a sound source and to separate out multiple sources located at different positions, as well as to the amount of information transfer that can be achieved in the identification of source position. Response bias refers to the average differences between perceived source position (as measured by the mean of the listener's objective responses) and the actual source position.

Ideally, the alterations introduced to improve localization performance should improve resolution without increasing response bias. However, this is generally not possible: alterations introduced to improve resolution will generally increase response bias. On the other hand, in many cases, the response bias can be reduced, if not eliminated, by appropriate experience with the altered cues (sensorimotor adaptation). Thus, the challenge in this work is to develop alterations or transformations of localization cues that (a) can result in substantially improved resolution and (b) produce response biases that can easily be reduced to manageable size by appropriate training or exposure experience. Furthermore, since in most practical applications it will be necessary for the human user to switch back and forth between the system with the altered cues and the normal environment, it is important that the adaption required to eliminate the response bias with the altered cue system be easily reversible (in the sense of not causing the listener to be seriously maladjusted to the normal environment).

The overall, long-term objectives of this work are to (1) determine, understand, and model the perceptual effects of altered auditory localization cues, and (2) design, construct, and evaluate cue alterations that can be used to improve performance of human-machine interfaces in virtual-environment and teleoperator systems. This research differs from most other research concerned with spatial localization in auditory displays in its concern with supernormal performance and adaptation to altered cues. It differs from most other research on perceptual rearrangement and adaptation in its focus on improving performance and its concern with resolution as well as response bias.

Because of equipment limitations, a greater portion of resources was devoted to equipment issues than originally planned. For the same reason, the cue alterations studied experimentally have been restricted to ones in which the set of normal acoustic cues are retained, but the mapping between this set of cues and spatial location is altered. In particular, attention has been focussed on how performance in the identification of sound source azimuth is affected by transforming the

azimuthal parameter θ in such a way that resolution is increased in the neighborhood of $\theta = 0^0$ (i.e., straight ahead) and decreased in the neighborhoods of $\theta = \pm 90^0$ (i.e., off to the sides). In other words, a transformation has been selected that increases the extent to which the dependence of azimuthal resolution on azimuth evidences the characteristics of an "acoustic fovea".

In a variety of experiments performed with this transformation, it has been shown that resolution changes in roughly the anticipated manner, that substantial response bias of the type expected occurs when this transformation is introduced, and that this response bias can be at least partially eliminated by appropriate training procedures. Although these general results appear relatively robust (in the sense of being roughly invariant over a variety of changes in detailed experimental procedure), two features of the results remain puzzling. First, in most cases, the improvement in resolution associated with the introduction of the altered cues tended to diminish somewhat as exposure to the altered cues increased and the subject adapted to the alteration (i.e., the response bias decreased). Second, under one condition tested (in which the subject was deprived of all visual input), no adaptation took place. Clearly, a full understanding of both these features is essential both from the point of view of basic research and from the applications viewpoint.

In addition to this series of experiments, all of which were performed in a hybrid environment consisting of a virtual auditory environment and a real visual environment, significant effort was devoted to the development and acquisition of improved facilities. Included in this category are the construction of a head-mounted display and a pseudophone, as well as an experimental system for doing hand-pointing experiments with a virtual sound source located in the hand. Also, substantial progress was made in the development of an inertial tracker for monitoring head or hand movement that is expected to be superior to most current trackers in minimizing delay and maximizing work space (this project has been partially supported by NASA contract NCC2-771). Further activities oriented towards the development of improved facilities focussed on the development and/or acquisition of alternate real-time cue-synthesis systems (e.g., acquiring a simplified analog cue-synthesis system for comparison purposes; and helping to formulate specifications for an improved digital cue-synthesis system).

Finally, in addition to expanding and refining the conceptual framework for research in this area (Durlach, 1991; Durlach, Rigopulos, Pang, Woods, Kulkarni, Colburn, and Wenzel, 1992; Durlach, Shinn-Cunningham, and Held, 1993), theoretical computations were made to check the extent to which it is appropriate to simulate an enlarged head, which is one of the more obvious ways to magnify localization cues in both azimuth and elevation, by means of frequency scaling (Rabinowitz, Maxwell, Shao, and Wei, 1993).

Final Report

The general objectives of our initial grant on Super Auditory Localization were to "determine, understand, and model the perceptual effects of altered localization cues." We had initially intended to conduct this work using a virtual-environment (VE) system for visual as well as auditory stimulation, and to include examination of a wide variety of transformations (rotations, scalings, filterings, asymmetries, exponentiations).

As will be seen in the following discussion, the work we have completed under this initial grant has not achieved these general objectives. Furthermore, our work was conducted using a hybrid VE in which the acoustical stimulation was virtual but the visual stimulation was real, we studied only one transformation, and we made no effort to measure our own HRTFs. The decision to use available HRTFs rather than to construct our own was based on the realization that, at least for our purposes, such work would have a relatively low payoff-to-effort ratio compared to other work that needed to be done. Both the hybrid VE and the transformation used are described in Sec. B below.

The gaps between our stated objectives and our actual accomplishments are the result of a number of factors. The first and most important is that the total funding we have received constitutes only a small fraction of the funding that we requested in order to achieve the above-stated goals. Whereas our initial grant proposal was for five years (Nov. 1, 1989 - Oct. 31, 1994) and totalled roughly $1,501,000, the total amount of funds that we have actually received to date for this project is $595,000 ($547,000 from AFOSR and $48,000 from NASA). (All figures are Total Costs, not Direct Costs). Secondary factors include (1) the complexity of the subject addressed, (2) the relatively high cost and limited performance of the VE equipment that was available during the working period of this grant, and (3) the departure from MIT during this grant of a key research scientist assigned to this grant (X. D. Pang, for personal reasons). In light of these factors, we believe that our progress, discussed in detail in the following subsections, has been substantial.

A. Equipment Issues

The work originally envisioned on Super Auditory Localization for Improved Human-Machine Interfaces depended strongly on the availability of adequate technology for the presentation and control of acoustic and visual stimuli. Because the technology available proved to be less than adequate, a number of our research goals were scaled back or altered to fit the capabilities of the devices available. In addition, the development of improved equipment became a goal of the project.

One of the original objectives of the project was to investigate the use of auditory localization cues that exceeded the range of normal cues (e.g., interaural time differences that exceeded those that occur naturally). Unfortunately, the Convolvotron (the special-purpose auditory spatialization system used to synthesize localization cues in our experiments) was designed to present normal localization cues and was found to be incapable of presenting localization cues outside the normal range. Although the hardware in the Convolvotron is capable of generating abnormally large interaural time differences for a single source in real-time, it cannot do so for four sources

simultaneously. Even making use of the Convolvotron for a single source with abnormally large interaural differences proved impossible due to software constraints. Furthermore, the Convolvotron can store HRTFs for only a small number of source positions and performs a spectral interpolation to simulate source positions between these stored locations. Although the possibility of significant interpolation error was a troubling (but unavoidable) problem even for the use of normal HRTFs, the errors introduced by interpolation of super-normal HRTFs would be even larger. Consequently, HRTFs containing larger-than-normal localization cues were not used with the existing Convolvotron. Because of these limitations with the Convolvotron, the acoustical localization cues for the reported experiments were drawn from the pool of normal acoustical cues (which were stored in the Convolvotron), and cue alterations were achieved by changing the mapping between these cues and the direction of the source relative to the head.

In addition to restricting the magnitude of localization cues simulated, the Convolvotron /head-tracking acoustic VE suffered from time delays and other distortions (e.g., the distortions induced by spatial interpolation of HRTFs discussed above). A great deal of this delay was the result of the Bird™ tracker employed. This tracker, although state-of-the-art when purchased, suffers from time delays on the order of tens of milliseconds. In order to test the importance of these effects, alternate synthesis methods were developed.

Fig. 1. The M.I.T. pseudophone configured to present supernormal interaural delays.

A second acoustic synthesis device, based on the system described in Loomis, Hebert, and Cincinelli (1990), was procured in order to further test the effects of artifacts associated with the use of the Convolvotron (as well as to explore the use of simplified cues). This second system uses highly simplified interaural level and monaural spectral cues. Since it is an analog system, interpolation of cues is not necessary with this device, as it is with the Convolvotron. Experiments with this device are currently under way; however, results are not yet available for inclusion in this report.

A pseudophone was designed and built at M.I.T., and will be used in future experiments on auditory adaptation (see Fig. 1). This electrcnic device employs microphones that can be located at various points relative to the head and that are connected to headphones worn by the subject. The pseudophone will allow presentation of unnaturally large interaural differences (in amplitude as well as time, although Fig. 1 shows the system configured solely to increase interaural time delay) which are perfectly correlated with the wearer's movements with essentially no delay between head movement and change in stimulus characteristics.. Also, background sounds will go through the same transformation as the intended targets since the auditory rearrangement depends upon the physical geometry of the microphones rather than synthesis of acoustic cues by signal processing methods. Attenuation of natural, unprocessed sounds is achieved by the use of insert earphones and acoustic muffs.

The time delays and small working volume associated with existing tracking systems inspired the development of an inertial tracking system in addition to development of the pseudophone. This tracker[1] will provide a large working volume, increased resolution, and better dynamic performance than existing tracking devices. A prototype inertial tracker for the three degrees of freedom associated with head orientation has been tested and will be ready for use in the near future.

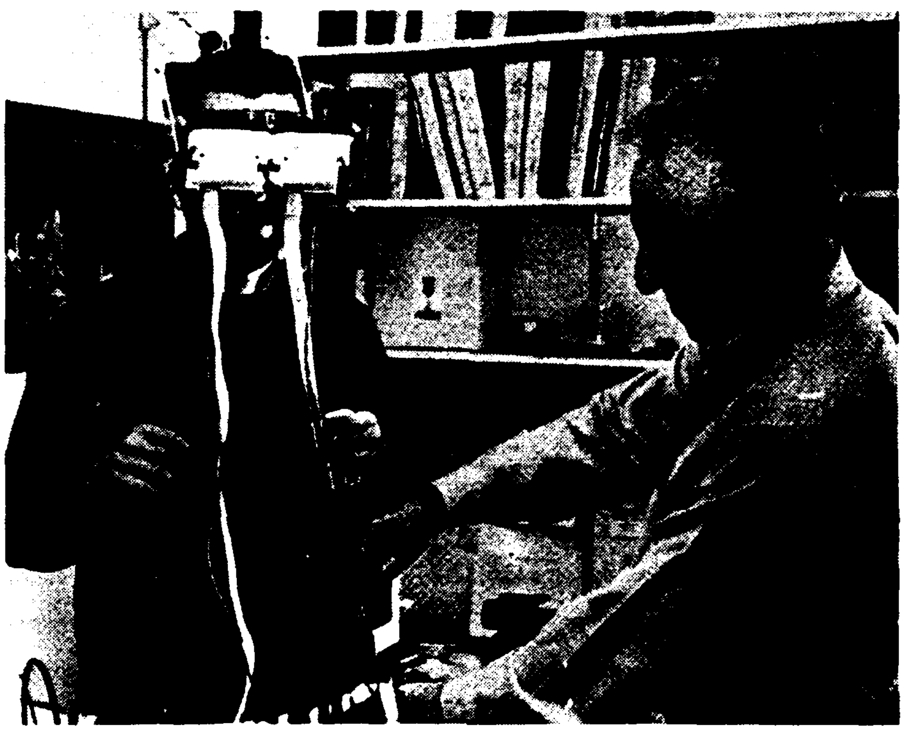

Fig. 2. The M.I.T. head-mounted display (HMD).

Finally, development of a visual virtual environment (VE) was originally undertaken to provide more flexible control of visual stimuli in the current project. A stereo head-mounted display (HMD) was developed in-house (see Fig. 2). This system, built from commercially available

[1]Development of the inertial tracker is being partially supported by NASA Contract No. NCC 2-771.

components, proved compact, lightweight, and portable. The HMD is completely untethered so that subjects can walk around freely when wearing it. The task of integrating the HMD with a graphics machine and the existing auditory VE in order to synthesize visual stimuli proved to be much more costly in both time and effort than was originally anticipated. While some progress on the development of a visual VE was made in the first year of the project, these efforts were put off in order to concentrate more fully on adaptation experiments that could be performed with the hybrid environment. The development of a visual VE is continuing under the sponsorship of contract N61339-93-C-0055 from the Navy.

B. Experimental Work

Adaptation to altered auditory localization cues was investigated by presenting simulated acoustic cues and real visual cues. Acoustic sources were "spatialized" by the Convolvotron, the special-purpose signal-processing system made by Crystal River Engineering and discussed above. The Convolvotron takes as inputs the source signal to be spatialized and the instantaneous position of the source relative to the listener's head and generates the binaural signals appropriate for a source from the specified position. In our system, the relative source position was calculated by a PC from the absolute position of the source to be simulated and the instantaneous orientation of the listener's head (reported to the PC by the Bird, a commercial head-tracking system).

This auditory virtual environment was used to simulate sources from one of thirteen positions around the listener at 0 degrees elevation, from -60 to +60 degrees in azimuth. These positions were indicated visually by a 3-foot-diameter arc of lights, which were clearly labelled (1 to 13) from left to right. These lights constituted our "real" visual display and were used to present visual spatial information about the simulated auditory sources presented to our subjects.

Auditory localization cues were transformed in this project by remapping the relationship between source position and Head Related Transfer Functions (or HRTFs). Using this approach, one defines a transformation

$$\theta' = f(\theta,\phi) \qquad \phi' = g(\theta,\phi)$$

and then defines

$$\begin{aligned} S'_L(\omega,\theta,\phi) &= S_L(\omega,\theta',\phi') = S_L[\omega, f(\theta,\phi), g(\theta,\phi)] \\ S'_R(\omega,\theta,\phi) &= S_R(\omega,\theta',\phi') = S_R[\omega, f(\theta,\phi), g(\theta,\phi)] \end{aligned} \tag{1}$$

where ω = frequency (in radians/sec), θ = azimuth, ϕ = elevation, $S_L(\omega, \theta, \phi)$ and $S_R(\omega, \theta, \phi)$ (which we refer to as the "*space filters*,") denote the complex transfer functions describing the filtering actions that occur as the sound propagates from the source to the left and right eardrums of the listener, and $S'_L(\omega, \theta, \phi)$ and $S'_R(\omega, \theta, \phi)$ are the transformed space filters.

In such a transformation, no new space filters are created; instead, the old space filters are reassigned to different angles. Moreover, they are assigned consistently with respect to the L-R variable so that not only are the same set of space filters used for each ear, but the same set of interaural ratios is preserved. In general, the use of such a remapping transformation will increase resolution in some regions of (θ, ϕ) space and decrease it in others.

As an illustration of this general class of transformations, confine attention to the horizontal

plane (i.e. assume $\phi = \phi' = 0$) and let

$$S'_L(\omega,\theta,0) = S_L(\omega,\theta',0)$$
$$S'_R(\omega,\theta,0) = S_R(\omega,\theta',0)\,, \tag{2}$$

where

$$\theta' = f_n(\theta) = \frac{1}{2}\tan^{-1}\left[\frac{2n\sin(2\theta)}{1-n^2+\left(1+n^2\right)\cos(2\theta)}\right]. \tag{3}$$

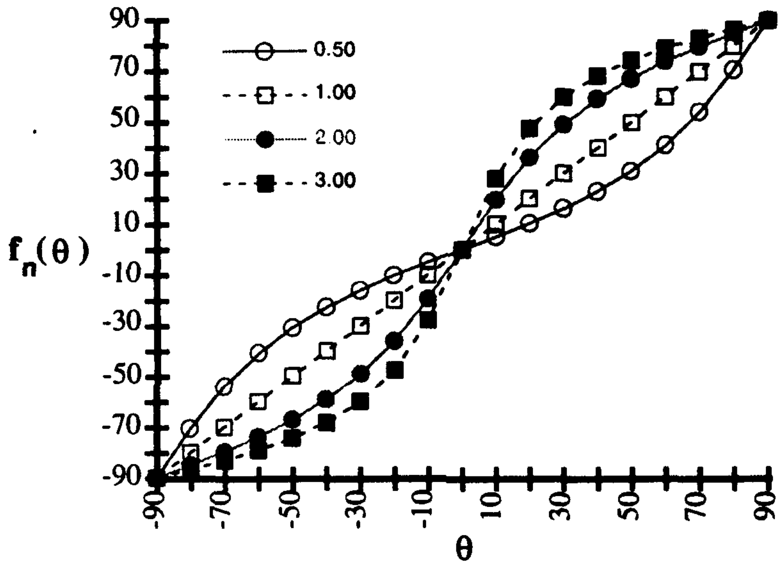

Fig. 3. A plot of the transformation specified by Eq. (3).

Pictures of this transformation are shown in Fig. 3 for the cases n = 1,2,3 and 1/2. For n>1 this transformation increases resolution in the neighborhood of $\theta = 0$ degrees and decreases it in the neighborhood of $\theta = +90$ degrees and -90 degrees. For n<1, the opposite occurs.

Transformations of this type may present less of a challenge to sensorimotor adaptation mechanisms than those considered in the previous two sections because no new auditory stimuli are introduced in these remapping transformations.

The only member of this family of functions that we have used to date is the function $f_3(\theta)$. With this function, source positions are displaced laterally relative to normal cues. The differences in localization cues for two sources in the frontal region (from -30 to +30 degrees in azimuth) are larger than normal with this remapping, while two locations off to the side give rise to more similar cues than are normally heard. With this transformation, subjects were expected to show better than normal resolution in the front and reduced resolution on the side, creating an enhanced "acoustic fovea" in which super auditory localization could occur. In addition to affecting resolution, however, this transformation was also expected to cause a bias whereby sources were perceived

farther off-center than were their actual locations. The main questions of the study were whether (1) bias could be overcome by subjects over time, so that they interpreted the new acoustic mapping of source position accurately, and (2) resolution was enhanced as expected in the "acoustic fovea". In all of our experimental work to date, attention has been focussed on the identification of source azimuth.

B-1. Experiment A

The basic experimental protocol consisted of a sequence of interleaved training and test runs. Each test run in the sequence consisted of 26 trials of a 13-alternative angle identification experiment. Test stimuli consisted of a 500 ms long click-train from one of 13 azimuthal positions separated by 10 degrees (ranging from -60 to +60 degrees). These positions corresponded to the positions of the lights, which were clearly numbered from left to right in an arc around the subject. Subjects had to face forward during each test stimulus or the trial was discarded. No correct-answer feedback was given and the lights were not used during the test runs. After each source was presented, the subject entered the number of the source position on a laptop keyboard.

During training runs, the subject was asked to track the source (whose position was chosen randomly for each trial from the set of 13 positions) by turning to point his/her nose to the correct location. During training, the light at the simulated acoustic location was turned on simultaneously with the acoustic source. In this manner, the subject became familiar with the mapping between source position, acoustic cues, and head orientation.

Each session (which lasted roughly 1.5 hrs) in this basic protocol consisted of the following sequence of test and training runs:

Test using normal cues	(1n)
Train using normal cues	
Test using normal cues	(2n)
- 5 minute break -	
Test using altered cues	(1a)
Train using altered cues	
Test using altered cues	(2a)
Train using altered cues	
Test using altered cues	(3a)
Train using altered cues	
Test using altered cues	(4a)
- 5 minute break -	
Test using altered cues	(5a)
Train using altered cues	
Test using normal cues	(3n)
Train using normal cues	
Test using normal cues	(4n)
Train using normal cues	
Test using normal cues	(5n)

Test Runs 1n, 1a, 5a, and 3n were analyzed in order to investigate how performance changed

over the course of each session. Run 1n provided a control against which other runs could be compared. Run 1a provided a measure of the immediate effect of the transformed cues. Any decrease in effect was found by comparing Runs 1a and 5a. Finally, Run 3n showed any negative after-effects due to exposure to the altered cues. The training and testing runs performed after Run 3n were included in order to help the subject re-adapt to normal cues. No special attention was given in these preliminary experiments to the issues of conditional or dual adaptation (e.g., Welch, 1978; Welch et al., 1993).

Using this paradigm, each of four subjects completed 8 identical sessions. Performance did not change significantly from the first to final session.

A couple of different data processing schemes were investigated. In one method based on standard psychophysical analysis methods (e.g. Durlach and Braida, 1969), the confusion matrix (matrix whose entry i, j corresponded to the number of responses i given when position j was presented) was analyzed for each subject and run, with multiple sessions combined within each such matrix (on the whole, we found comparatively little variation across sessions). With this approach, each source presentation was assumed to result in a stochastic decision variable with a Gaussian distribution along some internal decision axis. The mean of the distribution was assumed to depend monotonically on the source position and the variance was assumed equal for all source positions. Further, the decision axis was assumed to be broken into 13 contiguous regions corresponding to the 13 possible responses. In this model, if the sample of the decision variable fell into region 'i', the subject would respond "i". With these assumptions, a gradient-descent numerical algorithm was implemented to find the estimates of means and variances that maximized the likelihood of observing the given confusion matrix. From these maximum likelihood estimates, the sensitivity d_i' (a measure of the ability of the subject to discriminate between source positions i and i + 1) and bias β_i (a measure of the perceptual bias when position i is presented) were derived. While theoretically elegant, the solutions found with this method proved to be overly sensitive to outliers in the responses and numerically unstable.

In the second method, which proved to be both simpler and more robust, the average response and the standard deviation in response was found for each of the 13 possible locations for Runs 1n, 1a, 5a, and 3n (averaged across 8 sessions) for each subject. These two statistics (average response and standard deviation in response) were then used to estimate both resolution and bias for each run during the course of a session. Resolution between adjacent pairs of positions was estimated as the difference in mean responses normalized by the average of the standard deviations for the two positions. Bias (which is traditionally used to measure adaptation) was estimated as the difference between mean response and correct response, normalized by the standard deviation for the position. These metrics were averaged across subjects to generate a concise summary of results for each run (as with the variation across sessions, the variation across subjects was found to be relatively modest). This second approach determines estimates of d_i' and β_i that approach the maximum likelihood estimates found in processing method 1 as the number of response categories increases.[2]

[2]For experiments C and D, which used a pointing rather than identification response method, this second processing scheme yields the maximum likelihood estimates of d_i' and β_i.

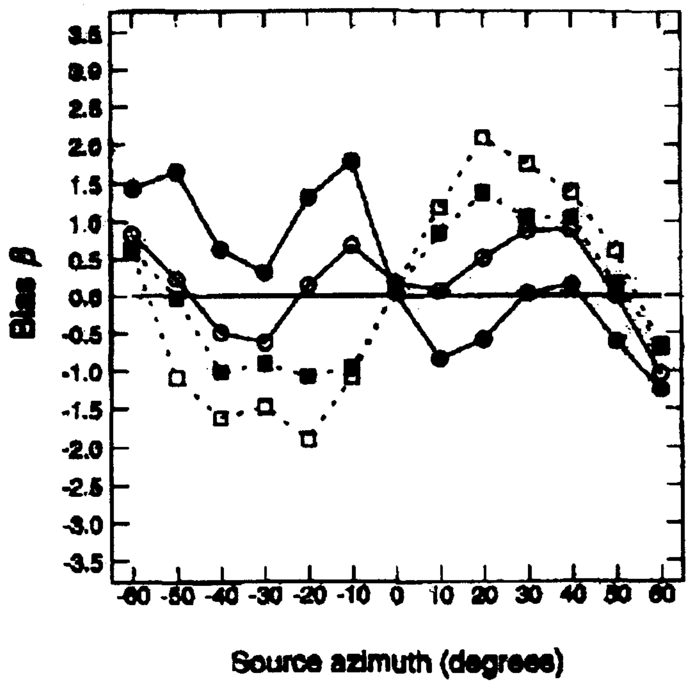

Fig. 4. Bias results for Experiment A. Normal cue tests are shown with circles, altered cue tests with squares. Open symbols represent tests prior to altered-cue exposure, filled symbols tests after exposure. The index i in d_i' and β_i has been omitted for simplicity.

Fig. 4 shows Experiment A bias results for runs 1n, 1a, 5a, and 3n as a function of source position (In this figure, as well as those that follow, the index i in d_i' and β_i has been omitted for simplicity). Normal-cue runs (1n and 3n) are plotted with circles; altered-cue runs (1a and 5a) with squares. The open symbols represent runs prior to altered cue training exposure (1n and 1a) while filled symbols correspond to the "adapted" results (5a and 3n). Results from Run 1n (open circles) showed some systematic biases, although these errors were significantly smaller than those found in other runs. In all bias results, there was an edge effect due to the experimental paradigm: since responses were limited to the 13 positions used, bias had to be positive (or zero) for the leftmost position (at -60 degrees azimuth) and negative (or zero) for the rightmost position (at +60 degrees azimuth). A strong bias occurred in Run 1a (open squares) in the direction predicted by the transformation and the aforementioned edge effect (subjects heard sources farther off-center than they were except for the leftmost and rightmost positions). Results from Run 5a (filled squares) showed a clear reduction in bias over the whole range of positions tested; however, this adaptation was not complete. Bias was reduced by roughly 30 percent with this experimental protocol. Finally, a negative after-effect is seen in the results from Run 3n (filled circles), where a strong bias was found in the direction opposite that induced by the altered cues.

Resolution results from Experiment A are shown in Fig. 5. Resolution for normal cue runs showed a systematic pattern (which may be due to systematic dependencies of the accuracy of the simulation on source position) which was consistent for pre- and post-exposure runs. Of more

interest is the comparison between normal- and altered-cue results. As expected with the transformation employed, resolution was enhanced for positions in the central region and decreased at the edges of the range.

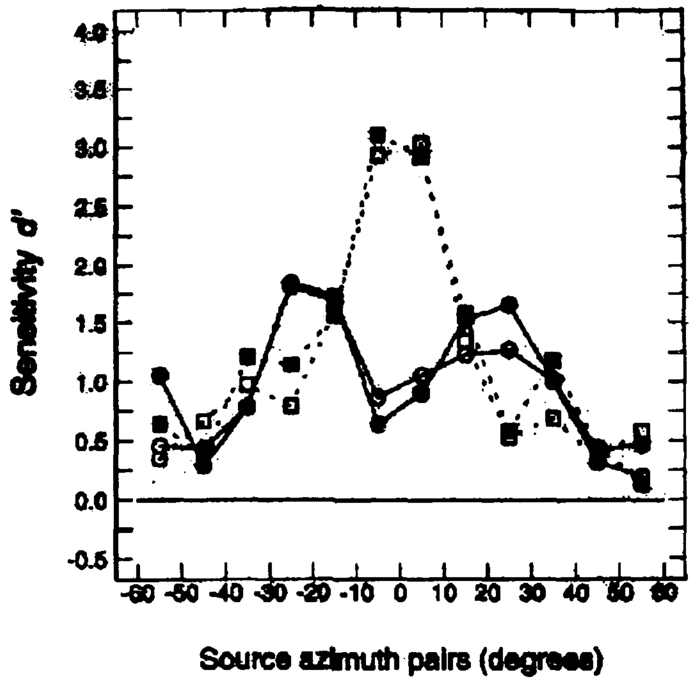

Fig. 5. Resolution results for Experiment A. See Fig. 4 caption.

B-2. Experiment B

Since only partial adaptation was found with the basic paradigm, a minor alteration in the stimuli was made to try to get more complete adaptation. Experiment B was identical to experiment A, except that a more complete "acoustic field" (analogous to the visual field discussed in Radeau and Bertelson, 1976) was simulated. Along with the click-train target, continuous sources were simulated outside of the range of target positions: a music source (Handel, 1740) from -90 degrees, and a voice (Auel, 1980) from 180 degrees. Since both -90 and 180 degrees are mapped to the same position with the remapping function $f_3(\theta)$ (see Eq. B-10), these "stable" sources were presented from roughly the same positions during both normal- and altered-cue runs. During training runs, the expectation that each source remained in one exo-centric position as subjects turned their heads provided additional information about the transformation. These sources were added in an effort to make the acoustic field more complex and rich in information, since some studies of visual or auditory illusory effects show a dependence on the number of sources visible or audible (e.g., Lackner, 1983).

Eight subjects performed Experiment B. Analysis yielded the bias results shown in Fig. 6 and the resolution results in Fig. 7. Bias results were very similar to those of Experiment A, with a strong immediate effect, a reduction of roughly 30 - 50 percent with exposure, and a strong negative after-effect.

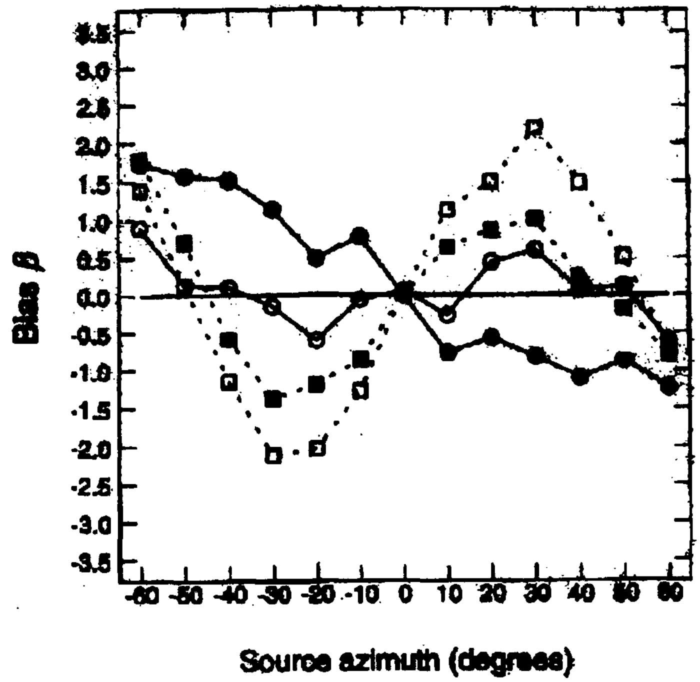

Fig. 6. Bias results for Experiment B. See Fig. 4 caption.

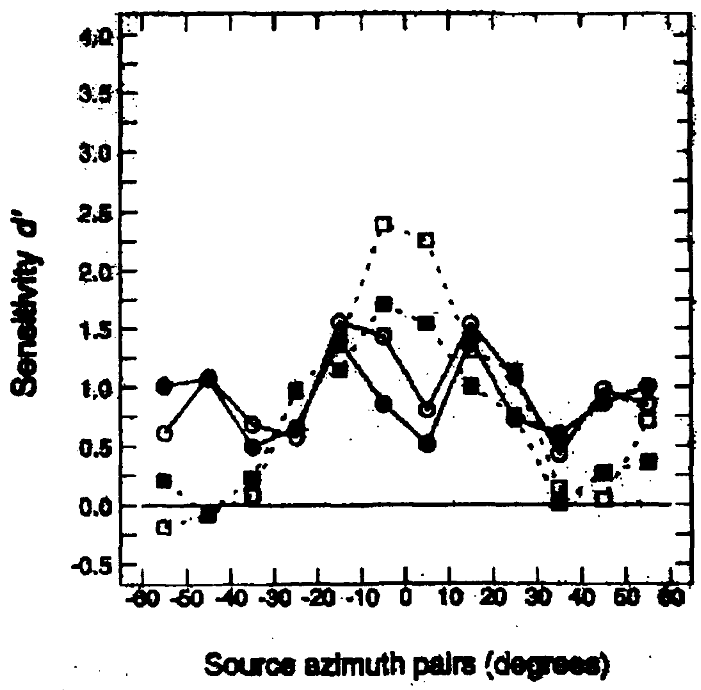

Fig. 7. Resolution results for Experiment B. See Fig. 4 caption.

The resolution results for normal cue runs in Experiment B showed the same systematic variation as those of Experiment A. Resolution for the first altered cue run in Experiment B was similar to that of the first experiment, although the increase in resolution for the center two pairs of positions was somewhat smaller than that seen in Experiment A. Of more interest, however, were the resolution results for the final altered cue run. In Experiment B, resolution appeared to decrease significantly for the center positions with exposure to the altered cues.

B-3. Experiment C

In Experiment C, blindfolds were used to investigate whether adaptation could occur in the absence of visual cues. Five blindfolded subjects performed 8 sessions of testing and training. Since subjects were blindfolded and could not accurately type responses, the identification response method was abandoned in favor of a pointing response: subjects were asked to turn their noses to point to the position of the click train after each presentation (subjects still had to face forward during each test stimulus or the trial was discarded). With the exception of the blindfold and the response method, Experiment C was identical to Experiment A.

Bias results from Experiment C (shown in Fig. 8) are strikingly different from those of the previous experiments. No reduction in bias occurred with exposure, nor was there any negative after-effect.

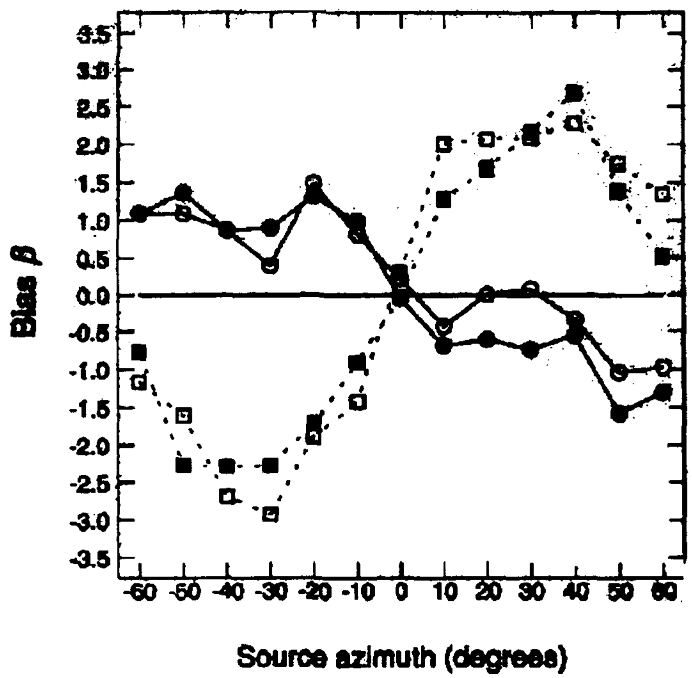

Fig. 8. Bias results for Experiment C. See Fig. 4 caption.

In addition to the clear lack of adaptation with the experimental paradigm of Experiment C, other differences of note occurred. The edge effects seen in the previous experiments were much

less pronounced. Subjects were told verbally that only positions from -60 to +60 degrees in azimuth would be presented and were shown the possible source locations on the labelled light-arc prior to putting on the blindfolds at the start of each session, yet they still consistently turned outside ‸e range of possible positions for altered cue sources at the edges of the azimuthal range. With ‸rmal cues, subjects showed a clear tendency to under-estimate the lateral position of the simulated sources, again in contrast to the previous experimental results. These differences are thought to be the result, at least in part, of the response method.

Resolution results for Experiment C are shown in Fig. 9. Resolution is somewhat enhanced in the central region (both before and after exposure) with altered cues, with the increase in resolution close to that seen in Experiment B. A slight decrease in resolution with exposure to the altered cues occurred in this experiment, but was not as pronounced as in Experiment B.

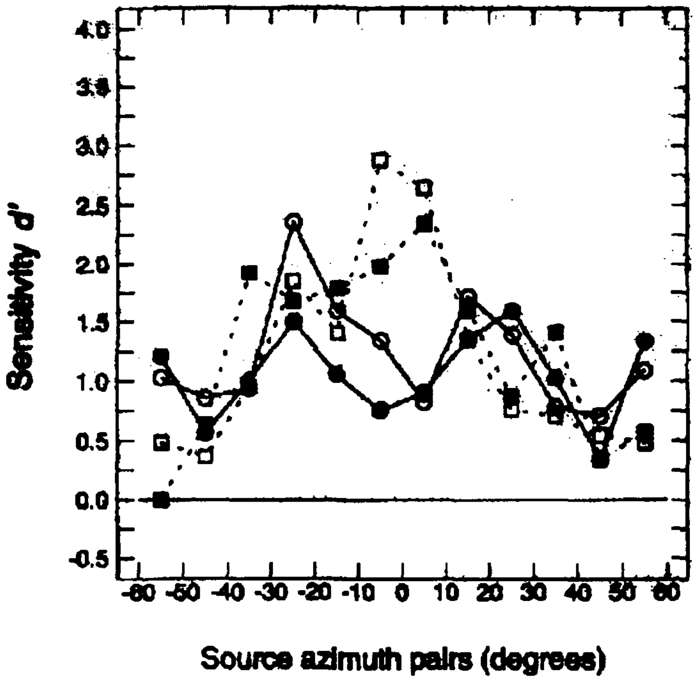

Fig. 9. Resolution results for Experiment C. See Fig. 4 caption.

B-4. Experiment D

Experiment D was performed to test whether the lack of adaptation in Experiment C was the result of the altered response method or the lack of visual stimuli. The experimental paradigm used in Experiment D was identical to that of Experiment C, except that subjects were not blindfolded. The visual scene in the room was thus available to the subjects in this experiment, and subjects were exposed to correlated light/sound sources during training. Unfortunately, time limited the number of sessions performed by the four subjects who performed Experiment D: 3 of the subjects finished 2 identical sessions each, while the fourth finished 3 sessions.

Bias results from Experiment D are shown in Fig. 10. These data are clearly much noisier than any of the previous results. This is to be expected, since at least 4 times as many points were

averaged in the previous results compared to those shown here. Although conclusions drawn from the results of Experiment D are tentative at best, there does seem to be adaptation occurring for the data from the left side of the source range (from -60 to 0 degrees in azimuth). These bias results are very similar to results from Experiments A and B. The usual strong immediate effect is reduced in these data by nearly 50 percent with exposure to the altered cues, while a negative after-effect also occurs. On the whole, the results for the right side of the source range are not systematic. Examination of the raw responses for source positions to the right uncovered a large number of outliers in the responses for this half of the data. Given the small number of points averaged for the plots in Fig. 10, these outliers had a huge effect on the results for positions to the right of center, so that any effects which may have occurred were obscured by the noise.

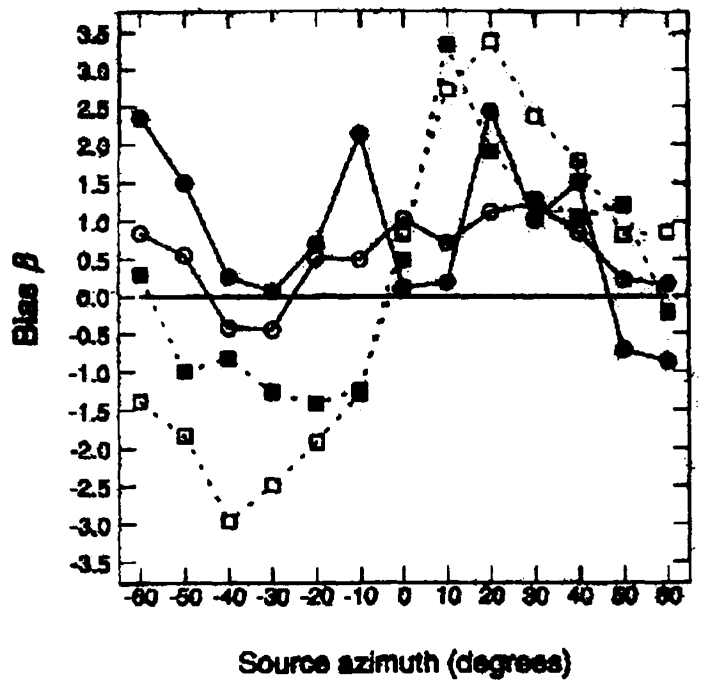

Fig. 10. Bias results for Experiment D. See Fig. 4 caption.

Estimates of resolution for Experiment D are shown in Fig. 11. Again, the small amount of averaging for this experiment makes strong conclusions difficult. Resolution at the central two positions is elevated for both runs using altered cues; however, the random fluctuations in the normal run resolution data are larger than this resolution increase.

The results of Experiment D tentatively point to the blindfolding of subjects as the significant change in experimental paradigm between Experiments A and B and Experiment C. Time prevented detailed exploration of the dependence of adaptation on vision; however, the importance of vision to auditory spatial adaptation is not surprising. A large number of studies (Warren and Pick, 1970; Canon, 1970; Pick, Warren, and Hay, 1969; Jones and Kabanoff, 1975; Mastroianni, 1982; Platt and Warren, 1972; Ryan and Schehr, 1941) implicate vision as uniquely important in spatial perception.

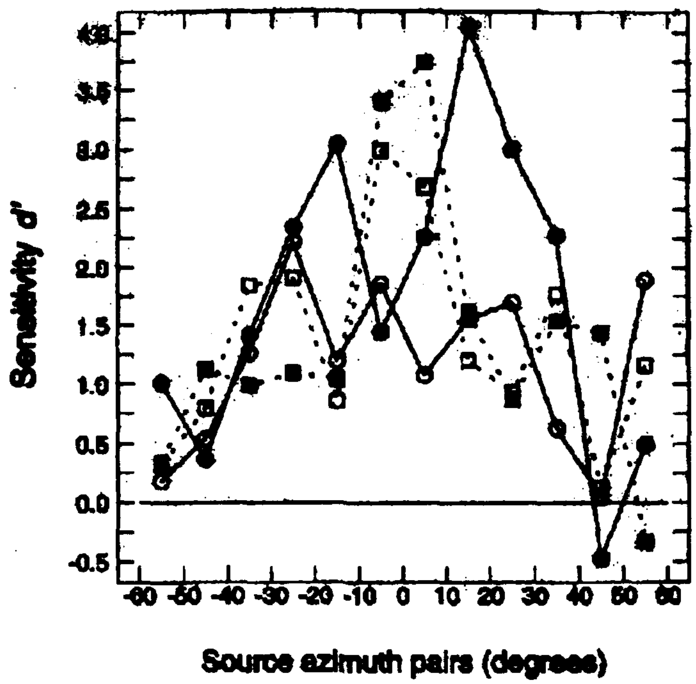

Fig. 11. Resolution results for Experiment D. See Fig. 4 caption.

B-5. Experiment E

The first four experiments were done in a manner consistent with most previous work on adaptation, by using a training procedure that involves both the sensory and motor systems. In the psychophysical literature, training is often accomplished with correct-answer feedback, which is strictly cognitive in nature, and without motor involvement. To see if similar adaptation results could be obtained using general psychophysical procedures, Experiment E was performed without any active training runs, but with correct-answer feedback given after each trial by flashing the light at the correct location after the subjects entered his/her response.

In contrast to Experiments A and B, subjects never were given auditory and visual stimuli simultaneously, although visual stimuli were presented following auditory stimuli from the same location. Also, subjects did not experience localization cues involving the entire sensorimotor loop, since only testing runs (during which subjects faced forward during each presentation) were employed in Experiment E. As in Experiments A and B, subjects entered their responses on a keyboard rather than using the head-pointing response method. Three sources were present during every run (as in Experiment B). In order to make the exposure times similar to those of the previous experiments, 40 test runs of 26 trials each were used in Experiment E. Each session of 40 test runs lasted between an hour and an hour and a half. The order of the runs was

2 tests with normal cues	(1n-2n)
8 tests with altered cues	(1a-8a)
- 5 minute break -	
22 tests with altered cues	(9a-30a)

- 5 minute break -

8 tests with normal cues. (3n-10n)

In order to reduce variability, pairs of runs were analyzed together for the five subjects who performed 8 sessions of Experiment E. Thus, Runs 1n and 2n were averaged together across 8 sessions for each subject to give the normal cue baseline of performance; Runs 1a and 2a were combined to examine the immediate effect of the transformation; Runs 29a and 30a were averaged to examine the decrease in effect; and Runs 3n and 4n gave a measure of negative after-effect.

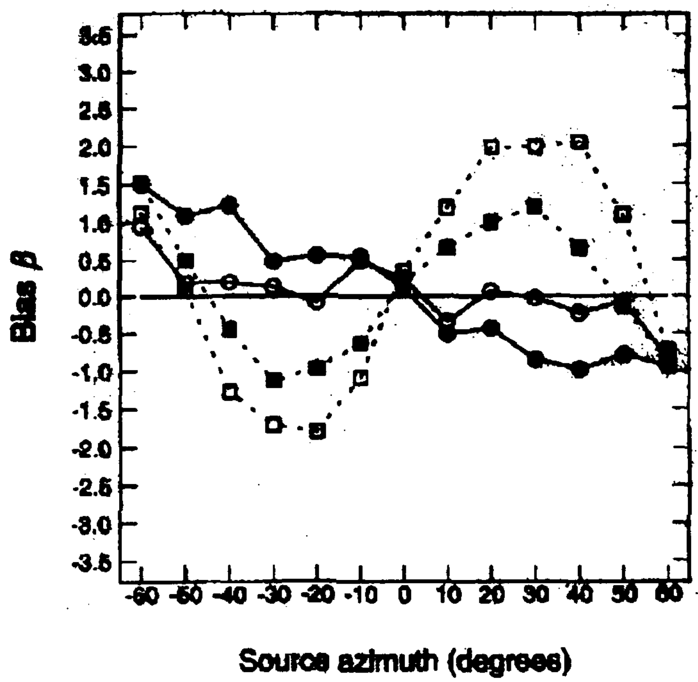

Fig. 12. Bias results for Experiment E. See Fig. 4 caption.

Bias results from Experiment E (shown in Fig. 12) closely resemble the results of Experiments A and B. An immediate effect is seen which follows predictions for the transformation and response method employed. The bias is reduced by about 30 percent with repeated exposure to the transformation (by correct-answer feedback in this case). When normal cues are tested following the altered cue tests, subjects show a strong negative after-effect.

Resolution results (shown in Fig. 13) are very similar to those of Experiment B. Resolution is enhanced in the first altered cue test for the center positions; however, this increase is reduced by the last altered cue tests. As in Experiment B (and unlike Experiment A), an ongoing music source was present from -90 degrees and a voice source from 180 degrees.

B-6. Experiment F

The decrease in altered-cue resolution with time seen in all experiments but A, although in many cases of small magnitude, was surprising. Since peripheral resolution for the center positions was enhanced with the altered cues, it is reasonable to assume that the decrease in

resolution over time must come from central mechanisms. Furthermore, if such were the case, then simplification of the task might eliminate the decrease over time. With this in mind, Experiment F was performed using only the center seven locations. This change in the stimulus set simplified the task not only by decreasing the number of stimuli, but also by restricting the stimuli to a region where resolution always increased or remained unchanged (so that the resolution change was no longer non-monotonic). Experiment F was identical to Experiment E (with 2 continuous sources along with a target click train), except that only the seven center source positions were used.

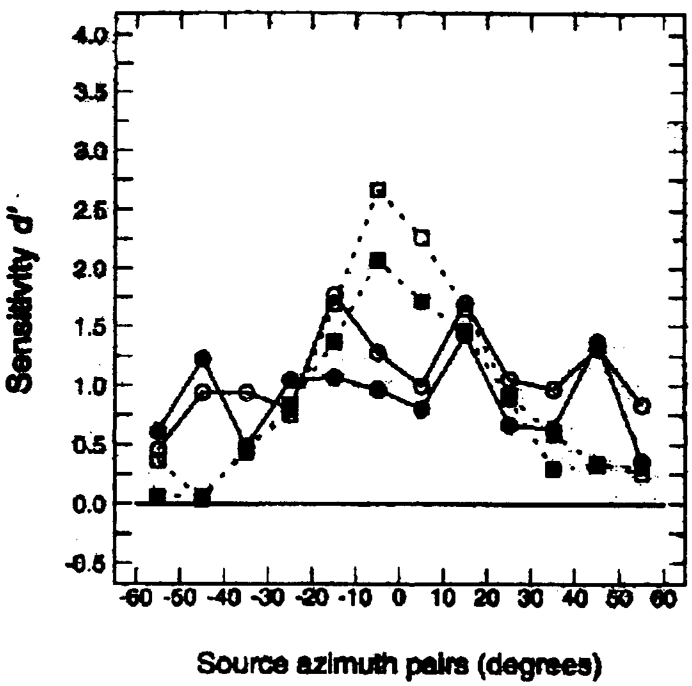

Fig. 13. Resolution results for Experiment E. See Fig. 4 caption.

Bias results for Experiment F, shown in Fig. 14, show the expected pattern of results. While the edge effect for Experiment F reduces the size of the immediate bias measured with the 7-alternative identification task, the bias is reduced by over 50 percent by the end of the altered-cue exposure period. The negative after-effect in Experiment F is at least as strong as was seen in previous experiments.

Resolution results are seen in Fig. 15. The results clearly show an increase in resolution for the center positions. Most importantly, resolution remains enhanced throughout the altered-cue exposure time.

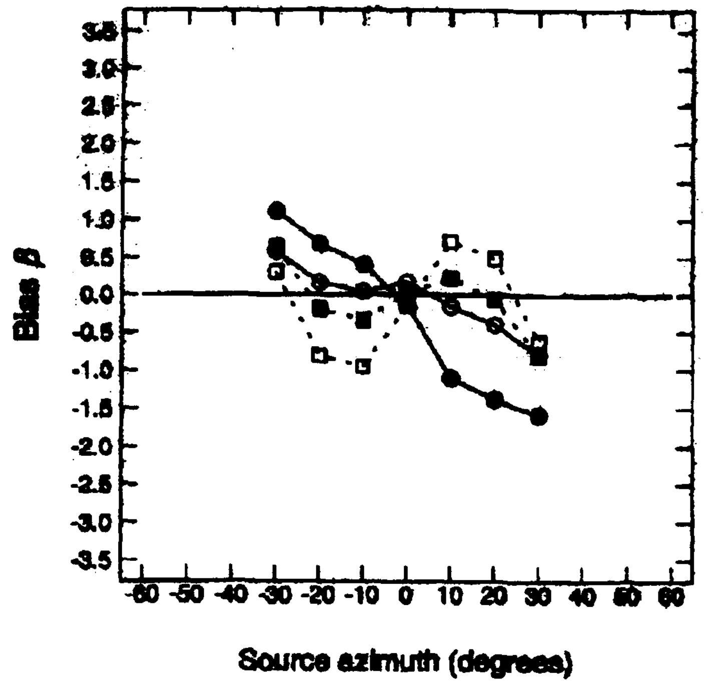

Fig. 14. Bias results for Experiment F. See Fig. 4 caption.

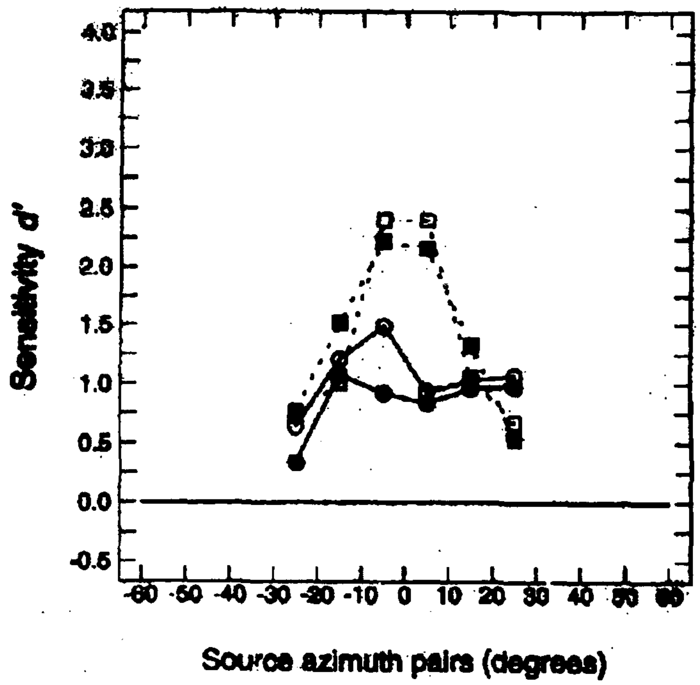

Fig. 15. Resolution results for Experiment F. See Fig. 4 caption.

B-7. Hand-Pointing Experiments

A number of the previous experiments investigating auditory adaptation (e.g., Mikaelian and associates, Freedman and associates, and Kalil) employed paradigms in which a sound source was held in the hand of the subject. In these experiments, cues were altered with a pseudophone while the subject made pointing responses with the hand holding one source to match the position of a target source. In order to determine whether shifting the paradigm in this manner would alter our results, another testing paradigm was developed which used the Convolvotron in conjunction with a tracker worn on the hand. In these experiments, subjects were seated with their heads held stationary in a head rest. A target source was presented at one of ten positions around the subject, and the subject was asked to make a ballistic pointing movement with his right hand to match the azimuthal position of the target. When the hand reached the end of its trajectory, a source simulated at the hand was turned on, and subjects heard the extent of their pointing error. This experimental paradigm was perfected in a series of pilot tests, and is currently being used to test a set of six subjects. Results from these tests are not yet completed, and could not be included in the current report; however, pilot results indicate that this testing paradigm will yield results similar to those that we have already reported using other methods.

C. Computations Concerning the Use of Frequency-Scaling to Simulate an Enlarged Head

In examining ways in which supernormal localization cues could be produced, the idea of generating HRTFs from larger-than-normal heads was considered. Large-head HRTFs could be tested with equipment and experimental paradigms similar to those used in the previous experiments, once the large-head HRTFs were produced. One way of generating large-head HRTFs would be to build a physical model of a larger-than-normal head, and to empirically measure the resultant cues. This method is not only very time consuming, but also inflexible, since for every new head-size to be tested, the whole procedure would have to be repeated.

An alternate approach would derive large-head HRTFs from empirically measured, normal HRTFs. One method for doing this is to use frequency scaling. In anticipation of employing this method, the theoretical effects of frequency-scaling HRTFs to approximate a larger than normal head were investigated and reported in Rabinowitz, Maxwell, Shao, and Wei, (1993). In this work, it was shown that frequency-scaling normal HRTFs will produce results very similar to HRTFs from larger than normal heads, provided the sources to be simulated are relatively far from the listener.

D. Comments

In all six experiments (A-F), introduction of the transformation $f_n(\theta)$ produced the anticipated changes in resolution and, in particular, increased resolution in the center of the field. Furthermore, most of the experiments (all but C) showed clear evidence of adaptation. Not only did the subjects in these experiments show a reduction in bias (and localization error) with exposure to the altered cues, but also an increase in bias (and localization error) in the opposite direction when tested with normal cues following altered-cue exposure (the negative after-effect).

Of particular interest in the context of classical adaptation work, adaptation was found to occur, with essentially comparable strength, without involvement of the sensorimotor system in the adaptation process: in both experiments E and F, the feedback was purely cognitive. This result contrasts with previous work (e.g., by Held, by Mikaelian and associates, by Freedman and associates, and by Kalil) which focussed strongly on the importance of sensorimotor involvement.

Independent of the issues of sensorimotor involvement, two results complicate the above simplified picture. First, no adaptation occurred in the experiment in which the subjects were blindfolded (Exp. C). Second, in Exps. B, C, D, and E, there appeared to be a tendency for the enhanced resolution to decrease with increased exposure to the altered cues and (except for Exp. C) the occurrence of adaptation.

We currently have three distinct ideas about the observed decrease in resolution enhancement. The first, already mentioned above in our discussion of the motivation for performing Exp. F, is that the decrease is associated with limitations of central processing and that the decrease will tend to disappear if the central processing load is decreased. This notion gains some support from the results of Exp. F.

The second idea is that the apparent decrease in resolution is an artifact resulting from the pooling of data over runs in which the response criteria are being altered by the subject in association with the decrease in response bias (i.e., with the adaptation to the altered cues). If this hypothesis were correct, one would expect the resolution enhancement to increase again after the criteria stabilized. This hypothesis is appealing but inconsistent with two observations: (a) In Exps. A and F, where substantial adaptation was seen, the loss in resolution enhancement was negligible; (b) In Exp. C, where no adaptation took place, a slight loss in resolution enhancement was seen. It should be noted, however, that the strength of these two observations (i.e., their statistical significance) has not yet been estimated.

Finally, the third idea is that the implicit assumption in our work that resolution and bias are independent is false, and that the criteria placement that corresponds to the elimination of bias leads to reduced resolution enhancement. In this third hypothesis, in contrast to the second, it is not the movement of the criteria from one set of positions on the decision axis to another that is the culprit, but the final location of these criteria on the decision axis. Clearly, further analysis and further experiments are required to identify the underlying cause of the reduced resolution enhancement (as well as the failure to adapt in Exp. C).

E. Publications, Talks, Meetings, and Patents

E-1. Publications

Durlach, N. I. (1991). "Auditory Localization in Teleoperator and Virtual Environment Systems: Ideas, Issues, and Problems," Perception, **20**, 543-554.

Durlach, N. I., Rigopulos, A., Pang, X. D., Woods, W. S., Kulkarni, A., Colburn, H. S., and Wenzel, E. M. (1992). "On the externalization of auditory images," Presence, **1**, 251-257.

Durlach, N. I., Shinn-Cunningham, B. G., and Held, R. M. (1993). "Super Auditory Localization. I. General Background," Presence, in press.

Rabinowitz, Maxwell, Shao, and Wei. (1993). "Sound localization cues for a magnified head:

Implications from sound diffraction about a rigid sphere," Presence, in press.

E-2. Talks

Durlach, N. I. (1991). "Sensing and Displaying Acoustic Information," ILP Symposium on Telerobotics, MIT, Oct. 29-30, 1993.

Durlach, N. I. (1991). "Super Auditory Localization for Improved Human-Machine Interfaces," DOD User-Computer Interaction Technical Group, San Antonio, TX, Nov. 5, 1991.

Durlach, N. I., Held, R. M., and Shinn-Cunningham, B. G. (1992). "Super Auditory Localization Displays," Society for Information Display International Symposium Digest of Technical Papers, vol. XXIII, 98-101.

Shinn-Cunningham, B. G., Durlach, N. I., and Held, R. (1992). "Adaptation to transformed auditory localization cues in a hybrid real/virtual environment," J. Acoust. Soc. Am., **92**, 2334.

Shinn-Cunningham, B. G. (1993). "Auditory virtual environments," talk presented at the M.I.T. Workshop on Space Life Sciences and Virtual Reality, Endicott House, Dedham, MA, 6 January 1993.

Shinn-Cunningham, B. G., Durlach, N. I., and Held, R. (1993). "Super Auditory Localization for improved human-machine interface," talk presented at the AFOSR Review of Research in Hearing, Fairborn, OH, June 1993.

A talk entitled "Auditory Displays and Localization" is being prepared for presentation at the Conference on Binaural and Spatial Hearing sponsored by the AFOSR and Armstrong Laboratory, WPAFB, September 9-12, 1993.

E-3. Meetings

Additional work connected with this grant has involved participation in meetings at government agencies (e.g., NASA and ONR) and participation in meetings of the Acoustical Society of America.

E-4. Patents

An invention disclosure has been submitted to M.I.T.'s Office of Technology Licensing for the work on an inertial tracking system. A patent may ensue for this tracker, which was supported both by this project and NASA contract NCC 2-771.

References

Auel, J. M. (1980) *Clan of the Cave Bear*, book on tape, Crown, New York.

Canon, L. K. (1970). "Intermodality inconsistency of input and directed attention as determinants of the nature of adaptation," J. Exp. Psych, **84**, 141-147.

Durlach, N.I. (1991). "Auditory Localization in Teleoperator and Virtual-Environment Systems: Ideas, Issues, and Problems", Perception, **20**, 543-554.

Durlach, N.I., and Braida, L.D. (1969). "Intensity Perception. I. Preliminary Theory of Intensity Resolution", J. Acoust. Soc. Am., **46**, 372-383.

Durlach, N.I., Rigopulos, A., Pang, X.D., Woods, W.S., Kulkarni, A., Colburn, H.S., and Wenzel, E.M. (1992). "On the Externalization of Auditory Images", Presence, **1**, 251-257.

Durlach, N. I., Shinn-Cunningham, B. G., and Held, R. M. (1993). "Super Auditory Localization," Presence, in press.

Freedman, S. J., and Gardos, G. (1965). "Compensation for auditory re-arrangement and transfer to hand-eye coordination" Paper read at MIT Conference on Adaptation, Cambridge, MA., June, 1965.

Freedman, S. F., Hall, S. B., and Rekosh, J. H. (1965). "Effects on hand-eye coordination of two different arm motions during adaptation to displaced vision," Percept. Mot. Skills, **20**, 1054-1056.

Freedman, S. J., and Rekosh, J. H. (1967). "Auditory effects of visual rearrangement," Paper read at MIT Conference on Adaptation, Cambridge, MA., June, 1965.

Freedman, S. J., and Stampfer, K. (1964a). "The effects of displaced ears on auditory localization," AFOSR Technical Report 64-0938.

Freedman, S. J., and Stampfer, K. (1964b). "Changes in auditory localization with displaced ears," Paper presented to The Psychonomic Society, Niagara Falls, Ontario, 9 October 1964.

Freedman, S. F., and Wilson, L. (1967). "Compensation for auditory rearrangement following exposure to auditory-tactile discordance," Percept. Mot. Skills, **25**, 961-966.

Freedman, S. J,. Wilson, L., and, Rekosh, J. H. (1967). "Compensation for auditory re-arrangement in hand-ear coordination" Percept. Mot. Skills, **24**, 1207-1210.

Freedman, S. J., and Zacks, J. L. (1964). "Effects of active and passive movement upon auditory function during prolonged atypical stimulation," Percept. Mot. Skills, **18**, 361-366.

Handel, F. (1740). "Concerti a due corni," Philips recording 7310 114 with the English Chamber Orchestra, directed by Raymond Leppard.

Held, R. M. (1955). "Shifts in binaural localization after prolonged exposure to atypical combinations of stimuli," Am. J. Psych., **68**, 526-548.

Held, R. M., and Hein, A. V. (1958). "Adaptation to disarranged hand-eye coordination contingent upon re-afferent stimulation," Percept. and Motor Skills, **8**, 87-90.

Jones, B., and Kabanoff, B. (1975). "Eye movements in auditory space perception," Percept. Psychophys., **17**, 241-245.

Kalil, R. and Freedman, S. J. (1967). "Compensation for auditory rearrangement in the absence of observer movement," Percept. Mot. Skills, **24**, 475-478.

Lackner, J. R. (1983). "Influence of posture on the spatial localization of sound," J. Aud. Eng. Soc., **31**, 226-233.

Loomis, J. M., Hebert, C. and Cincinelli, J. G. (1990). "Active localization of virtual sounds," J. Acoust. Soc. Am., **88**(4), 1757-1764.

Mastroianni, G. R. (1982). "The influence of eye movements and illumination on auditory localization," Percept. Psychophys., **31**, 581-584.

Mikaelian, H. H. (1969). "Adaptation to rearranged ear-hand coordination," Percept. Mot. Skills, **28**, 147-150.

Mikaelian, H. H. (1972). "Lack of bilateral generalization of adaptation to auditory rearrangement," Percept. Psychophys., **11**, 222-224.

Mikaelian, H. H. (1974). "Adaptation to displaced hearing: a nonproprioceptive change," J. Exp. Psych., **103**(2), 326-330.

Mikaelian, H. H., and Russotti, J. S. (1972). "Lack of bilateral generalization of adaptation to auditory rearrangement," Naval Submarine Medical Research Laboratory report number 715.

Pick, H. L., Warren, D. H., and Hay, J. C. (1969). "Sensory conflict in judgements of spatial direction," Percept. Psychophys., **6**, 203-205.

Platt, B. B., and Warren, D. H. (1972). "Auditory localization: the importance of eye movements and a textured visual environment," Percept. and Psychophys., **12**, 245-248.

Rabinowitz, W. M., Maxwell, J., Shao, Y., and Wei, M. (1993). "Sound localization cues from a magnified head: Implications from sound diffraction around a rigid sphere," submitted to Presence.

Radeau, M., and Bertelson, P. (1976). "The effect of a textured visual field on modality dominance in a ventriloquism situation," Percept. Psychophys., **20**, 227-235.

Rekosh, J. H., and Freedman, S. J. (1967). "Errors in auditory direction-finding after compensation for visual re-arrangement," Percept. Psychophys., **2**, 466-468.

Ryan, T. A. and Schehr, F. (1941). "The influence of eye movement and position on auditory localization," Am. J. Psych., **54**, 243-252.

Warren, D. H., and Pick, H. L. (1970). "Intermodality relations in localization in blind and sighted people," Percept. Psychophys., **8**, 430-432.

Welch, R.B. (1978). *Perceptual Modification: Adapting to Altered Sensory Environments*, Academic Press, New York.

Welch, R. B., Bridgeman, B., Anand S., and Browman K. E. (1993). "Alternating prism exposure causes dual adaptation and generalization to a novel displacement", *Percept. Psychophys.* **54**(2), 195-204.